

REMARKS

Applicants have now had an opportunity to carefully consider the Examiner's comments set forth in the Office Action of September 14, 2007.

Reconsideration of the Application is requested.

The Office Action

Claims 1, 20, 39, 58 and 81-84 stand provisionally rejected on the ground of non-statutory obviousness-type double patenting as being unpatentable over claims 1, 16 and 31-32 of co-pending Application No. 10/626,856.

Claims 1-5, 9-10, 14-24, 28-29, 33-43, 47-48, 52-62, 66-67, 71-76 and 81-88 stand rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 6,606,620 issued to Sundaresan et al. (hereinafter Sundaresan) in view of U.S. Patent No. 5,835,905 issued to Pirolli et al. (hereinafter Pirolli) and further in view of U.S. Patent No. 6,961,954 issued to Maybury et al. (hereinafter Maybury).

Claims 6-8, 25-27, 44-46 and 63-65 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Sundaresan in view of Pirolli and further in view of U.S. Patent Application No. 2004/006559 by Gange et al. (hereinafter Gange) and further in view of Maybury.

Claims 11-13, 30-32, 49-51 and 68-70 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Sundaresan in view of Pirolli and further in view of U.S. Patent Application No. 2004/002849 by Zhou (hereinafter Zhou) and further in view of Maybury.

Claims 77-80 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Brown, Ralf D. "Dynamic Stopwording for Story Link Detection" (hereinafter Brown) in view of U.S. Patent No. 6,012,073 issued to Arend et al. (hereinafter Arend).

The Double Patenting Rejections

Applicants again thank the Examiner for directing attention to the provisional double patenting rejection and note that once a patent issues with the asserted scope, Applicants will reevaluate the relevancy of the terminal disclaimer requirement, revise the scope of the pending application or take other appropriate action.

The §103 Art Rejections

The present Office Action rejects claims 1-5, 9-10, 14-24, 28-29, 33-43, 47-48, 52-62, 66-67, 71-76 and 81-88 under 35 U.S.C. §103(a) as being unpatentable over Sundaresan in view of Pirolli and further in view of Maybury. These rejections are respectfully traversed.

Claims 1, 3-19, and 85

Initially, referring to paragraph 6, page 3, of the Office Action, it states that Sundaresan teaches determining source-identified training stories (col. 3, lines 16-17, wherein "stories" means "documents"). Although the Sundareson reference clearly describes the use of training documents, the Office Action fails to show where the Sundaresan reference makes any reference to the use of source information for the training documents. For example, Sundaresan describes using a learning phase to develop models for classes with information it develops from the composite information gleaned from numerous training documents. And the Sundaresan reference further describes developing a structured vector model for each training document (col. 3, lines 14-18). However, contrary to any suggestion that Sundaresan makes use of source information in developing the vectors, the Abstract describes a structured vector model that allows like terms to be grouped together and dissimilar terms to be segregated based on their frequency and distribution within the sub-vectors of the structure vector, thus achieving context sensitivity. "Specifically, it develops a structured vector model for each training document. Then, within a given class of documents it adds and then normalizes the occurrences of terms" (Abstract). The Sundaresan reference appears to be silent on the subject of source information for the training documents, but instead describes only word or term-based vectors.

Further to the above, the Office Action admits that Sundaresan does not teach determining inter-story similarity vectors for at least one story-pair. The Office Action then states that Pirolli teaches determining inter-story similarity vectors, with reference to col. 7, lines 53-65. However, the vectors taught by Pirolli are clearly word-based and do not include source information. For example, Pirolli teaches that the "token information is then used to create a document vector, where each component of the

vector represents a word, step 403. Entries in the vector for a document indicate the presence or frequency of a word in the document. The steps 401-403 are repeated for each Web page in the Web locality. For each pair of pages, the dot product of these vectors is computed, step 404. The dot product which produces a similarity measure" (col. 7, lines 58-65, underlining added).

Contrariwise, claim 1 of the present application, as amended, recites limitations wherein the recited inter-story similarity vectors include two components: at least one inter-story similarity metric and at least one source-pair statistics. Fig. 6 of the present application shows the steps of determining source-pair similarity statistics which comprise a component of the inter-story similarity vectors. The process is further described in paragraphs 91-94 of the present application. It should be noted that in step S1030, the source pair statistics are determined based on the source characteristics of the stories in the source-pair. Specific source pair statistics are maintained for each identified source pair (par. 94).

Unlike the document vector of Pirolli, the recited source pair statistics are not document/story word or term based. As described in paragraphs 92-93, the source characteristics upon which the recited source pair statistics are based are associated with a source which may be a CNN, ABC, NBC, Aljazeera or CTV television broadcast, the text of a Reuters newswire service story, an article in the Wall Street Journal or any other known or later developed information source. The source characteristics associated with each source in a source-pair are used to select source-pair similarity statistics from the source hierarchy. The source hierarchy may be based on source characteristics such as source language, input mode and the like. An English radio broadcast captured using automatic speech recognition may be associated with an "English" language source characteristic and an "ASR" input mode source characteristic. A Chinese text translated into English may be associated with a "Chinese" source language characteristic and a "TEXT" input mode characteristic. The two stories thus form a story pair having "English:ASR" and "Chinese:TEXT" source pair characteristics.

Use of the above-described source-pair statistics as a component of similarity vectors, in combination with inter-story similarity vectors, as recited in claim 1, as

amended, are neither taught nor suggested in the cited references which describe only word/term-based vectors. Thus, the cited references, either individually or combined, do not teach the inter-story similarity vectors recited in claim 1, as amended. Applicants note that the limitations added to claim 1 are taken from the now-canceled dependent claim 2 and no new matter has been entered. Applicants also note that the Office Action cites col. 10, lines 15-17 of Sundaresan as teaching "determining at least one source-pair statistics" as now recited in claim 1, and previously recited in dependent claim 2. However, as described above, the statistics computed in the Sundaresan reference are term-based, and source information is not discussed. This is further made clear in lines 17-22 of col. 10: "The statistics are calculated by combining all the documents of a given type together in a meaningful fashion. In particular, the modeling sub-module 415 combines the individual vectors in the class by adding them together and normalizing the result. Term frequencies may be normalized at any level from the uppermost (document level) to the lowest sub-vector." The described step of adding individual vectors is merely adding previously determined term-based vectors and normalizing the results.

With reference now to a limitation for determining link label information as recited in claim 1 of the present application, the Office Action states that Sandaresan teaches "determining link label information for the at least one story-pair" (col. 9, lines 8-9). Applicants submit, however, that the classifier **10** referred to in the Sandaresan reference is only used to characterize the term frequency and distribution of the document in question and compare it to the known classes of documents (col. 9, lines 4-8, and Class 1 through class N in Fig. 5). Sundaresan does not teach or suggest that documents of the same class are necessarily linked.

However, the Office Action admits that Sundaresan does not teach the recited limitation "the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event", but the Office Action later states that the Maybury reference teaches the limitation with reference to col. 16, lines 31-33:

In addition, multiple story segment records 310, video theme records 312, video gist records 314, story summary records 315, and named entity records

316 result. Links are provided between the various records and the story segment to which they pertain, either directly or indirectly, to permit retrieval of related data.

The "records" Maybury describes above are in reference to database records and, therefore, the links clearly describe database links (as opposed to story link label information) so that the various records can be retrieved based on the link to the story segment to which they pertain. Note also that these links refer to a single story segment. The Maybury reference is not directed to linking different stories based being related to the same event, but contrariwise, is directed toward segmenting individual stories (Abstract). See also step (g) of claim 8: "(g) linking together a stored representation of the text data, summary data, and named entity data for the story segment." The Maybury reference appears to be silent on the subject of identifying stories linked by virtue of being related to the same event which is as one would expect since Maybury is directed to a system of automated segmentation of stories for presentation as broadcast news. Thus, even if the Maybury reference is combined with the Sandaresan reference, the combination would not produce the recited limitation of claim 1 of the present application.

For at least the reasons set forth above, it is submitted that claim 1 is distinguished over the references and in condition for allowance. As claims 3-19 and 85 depend from and further define claim 1, Applicants submit that these claims are also in condition for allowance.

Claims 20, 22-38, and 86

Claim 20 recites limitations similar to those of claim 1 as discussed above. Claim 20 has similarly been amended by incorporating the limitations of claim 21 which is canceled herein. Applicants submit therefore that, for the same reasons as set forth above, claim 20 is also distinguished over the references and in condition for allowance. As claims 22-38 and 86 depend from and further define claim 20, Applicants submit that these claims are also in condition for allowance.

Claims 39, 41-57, and 87

Claim 39 recites limitations similar to those of claim 1 as discussed above. Claim 39 has similarly been amended by incorporating the limitations of claim 40 which is canceled herein. Applicants submit therefore, for the same reasons as set forth above, that claim 39 is also distinguished over the references and in condition for allowance. As claims 41-57 and 87 depend from and further define claim 39, Applicants submit that these claims are also in condition for allowance.

Claims 58, 60-76, and 88

Claim 58 recites limitations similar to those of claim 1 as discussed above. Claim 58 has similarly been amended by incorporating the limitations of claim 59 which is canceled herein. Applicants submit therefore, for the same reasons as set forth above, that claim 58 is also distinguished over the references and in condition for allowance. As claims 60-76 and 88 depend from and further define claim 58, Applicants submit that these claims are also in condition for allowance.

Claims 81, 82, 83, and 84

Each of claims 81-84 recites limitations similar to those of claim 1 as discussed above. Each claim has been similarly amended by incorporating limitations similar to the limitations of now-canceled claim 2. Applicants submit therefore, for the same reasons as set forth above, that each of claims 81-84 is also distinguished over the references and in condition for allowance.

Claim 77

With reference now to claim 77, the Office Action states that Brown teaches the recited limitation of "a verified first source-mode transformation of the source-identified training corpus text from a first mode to a second mode." Firstly, the instant application and the language of claim 77 make it clear that the recited first source-mode transformation is a transformation of the text of the training corpus, e.g., transcription or translation as recited in the claim. That is to say, the content of the training corpus is transformed. The Office Action cites and characterizes the "single-pass incremental

clustering method" described by Brown as a transformation. However, the incremental clustering of Brown does not teach or fairly suggest any transformation of document or story content. The clustering is simply a method of organizing documents into clusters according to concepts described in Brown. For example, Brown describes "applying a penalty when the two documents under test are in different clusters" (page 1, col. 2, lines 39-40). The Office Action fails to show where the Brown reference teaches or fairly suggests a transformation of the text of the documents. Clearly, simply assigning a document to a specific cluster does not suggest modification of the document text such as the transcription and/or translation as recited in claim 77.

Further, with reference to the "second source-mode transformation" as recited in the claim, the Office Action is silent, and does not cite any reference as teaching a second source-mode transformation. Claim 77 of the present application, recites limitations wherein the same source-identified training corpus text undergoes two separate transformations, namely a first source-mode transformation and a second source-mode transformation. Even if the Office Action were correct in interpreting the clustering of whole documents as taught by Brown, the Office Action fails to show where two separate clusterings of the same source text are taught.

Continuing on the theme of two separate transformations of the same source-identified training corpus text as recited in claim 77, the claim further recites a limitation for "determining at least one transformation error associated with distribution differences between the first and second transformations." The Office Action cites page 2, col. 2, lines 4-6 of the Brown reference which reads: "A DET curve is generated by applying a continuously| variable threshold to the scores output by the system, arbitrarily setting the decision to YES for all scores above the threshold and to NO for scores below the threshold, and computing miss and false alarm rates for each value of the threshold." The underlined portion corresponds to the cited lines 4-6. However, the thresholds described by the Brown reference relate to different stories. For example, page 2, col. 2, lines 11-13 describe decisions made by thresholding on the reported score for the story pair. Not only does Brown's concept of thresholding not necessarily relate to errors, but, unlike the recited limitations of claim 77, it does not relate to different transformations of the same source. As a further example, on page 1, col. 2, lines 20-22, Brown describes

a "dual threshold is used to determine whether the computed cosine similarity indicates a linkage between the two stories." Again, Brown does not describe a threshold relating to different transformations of the same story.

For at least the reasons set forth above, it is submitted that claim 77 is distinguished over the references and in condition for allowance. As claims 78-80 depend from and further define claim 77, Applicants submit that these claims are also in condition for allowance.